Geostatistical investigation into the temporal evolution of spatial structure in a shallow water table

S. W. Lyon, J. Seibert, A. J. Lembo, M. T. Walter, T. S. Steenhuis

Hydrol. Earth Sys. Sci. Discuss., 2, 1683–1716, 2005
www.copernicus.org/EGU/hess/hessd/2/1683/
SRef-ID: 1812-2116/hessd/2005-2-1683
European Geosciences Union

Hydrology and Earth System Sciences Discussions

Papers published in *Hydrology and Earth System Sciences Discussions* are under open-access review for the journal *Hydrology and Earth System Sciences*

Geostatistical investigation into the temporal evolution of spatial structure in a shallow water table

S. W. Lyon[1], J. Seibert[2], A. J. Lembo[3], M. T. Walter[1], and T. S. Steenhuis[1]

[1]Biological and Environmental Engineering, Cornell University, USA
[2]Physical Geography and Quaternary Geology, Stockholm University, USA
[3]Crop and Soil Science, Cornell University, USA

Received: 1 August 2005 – Accepted: 15 August 2005 – Published: 29 August 2005

Correspondence to: T. S. Steenhuis (tss1@cornell.edu)

Abstract

Shallow water tables in the near-stream region often lead to saturated areas in catchments in humid climates. While these saturated areas are assumed to be of importance for issues such as non-point pollution sources, little is known about the spatial and temporal behavior of shallow water tables and the resulting saturated areas. In this study, geostatistical methods are employed demonstrating their utility in investigating the spatial and temporal variation of the shallow water table for the near-stream region. Event-based and seasonal changes in the spatial structure of the shallow water table, which directly influences surface saturation and runoff generation, can be identified and used in conjunction to characterize the hydrology of an area. This is accomplished through semivariogram analysis and indicator kriging to produce maps combining supplemental soft data (i.e., proxy information to the variable of interest) representing seasonal trends in the shallow water table with hard data (i.e., the actual measurements) that represent variation in the spatial structure of the shallow water table per rainfall event. The area used was a hillslope located in the Catskill Mountains region of New York State. The shallow water table was monitored for a 120 m×180 m near-stream region at 44 sampling locations on 15-min intervals. Outflow of the area was measured at the same time interval. These data were analyzed at a short time interval (15 min) and at a long time interval (months) to characterize the changes in the hydrology of the region. Indicator semivariograms based on transforming the depth to ground water table data into binary values (i.e., 1 if exceeding the time-variable median depth to water table and 0 if not) were created for both time interval lengths. When considering only the short time interval, the indicator semivariograms for spring when there is excess rainfall show high spatial structure with increased ranges during rain events with surface saturation. During the summer, when evaporation exceeds precipitation, the ranges of the indicator semivariograms decrease during rainfall events due to isolated responses in the water table. When summarized over a longer, monthly time interval, semivariograms exhibited higher sills and shorter ranges during spring and lower sills

and longer ranges during the summer. For this long time interval analysis, there was a good correlation between probability of exceeding the time-variable median water table and the soil topographical wetness index during the spring. Indicator kriging incorporating both the short and long time interval structure of the shallow water table (hard and soft data, respectively) provided more realistic maps that agreed better with actual observations then traditional hard data alone. This technique to represent both event-based and seasonal trends compensates for spatially sparse observations while incorporating physical hydrology of the hillslope to capture significant patterns in the shallow water table. Geostatistical analysis of the spatial and temporal evolution of the shallow water table gives information about the formation of saturated areas important in the understanding hydrological processes working at this and other hillslopes.

1. Introduction

Water tables occurring in nature are highly variable in both time and space. This variability creates difficulty in predicting how water tables respond to rainfall events and where saturated areas occur when the water table rises. This is troublesome because the position of the water table can determine which hydrologic pathways are active. Regions with high water tables can promote the occurrence of saturated areas leading to overland flow. How (i.e., through exfiltration, direct rainfall, or other pathways) and what (i.e., old or new water) water finds its way to these regions is not fully understood making the variability in physical patterns of saturated areas difficult to monitor and predict (McDonnell, 2003). These saturated areas are often typified by highly permeable surface layers underlain by highly impermeable subsurface layers at shallow depths. They act as runoff source areas causing runoff to be generated by rainfall amounts exceeding soil storage capacities. The difficulty in capturing the dynamics of these saturated source areas stems from the non-linear variability in both space and time exhibited among rain events and seasons. Due to this variability, researcher have coined the term variable source area (VSA) to describe these areas (e.g., Dunne and

Black, 1970; Hewlett and Hibbert, 1967; Dunne et al., 1975). While important in a pure hydrology perspective (i.e., predicting runoff amounts, peak timing in hydrographs), representing the spatial and temporal nature of VSAs is quintessential to modeling and managing contaminant flow pathways in natural environments. As observed by Grayson et al. (2002), there has been an increased focus in current research on spatial variability to account for where contaminants come from and where to invest financial resources to improve water quality. Although the concept has been around for well over a quarter of a century, it is obvious that the formation of VSAs and how they influence water quality is still a hot topic for hydrologist.

Repeatedly, the call for better distributed data to aid in understanding hydrological processes, especially for data to identify processes controlling the formation of VSAs, has gone out (Hillel, 1986; Klemeš, 1986; Hornberger and Boyer, 1995). New methods of collecting and interpreting spatially distributed data to characterize VSAs have become available. Snap shots of soil moisture using various remote sensing techniques (Choudhury, 1991; Engman, 1991; Blyth, 1993; Verhoest et al., 1998; Troch et al., 2000) and field measurements (Western and Grayson, 1998; Mohanty et al., 2000; Meyles et al., 2001; Walker et al., 2001; Wilson et al., 2004) have been used to locate regions concentrating water. These sampling techniques, however, may not be applicable for all field sites. For example, the extremely effective and increasingly popular technique incorporating time domain reflectrometry (TDR) sensors mounted to an all-terrain vehicle (Tyndale-Biscoe et al., 1998; Western and Grayson, 1998) is limited by field accessibility. This type of sampling may not be an option for field sites with large biota (e.g., trees, shrubs, corn), extreme geology (e.g., steep slopes, boulders, large gullys), or excessive amounts of surface water (e.g., ephemeral streams, saturated source areas). Satellite remote sensing techniques have their own difficulties such as signal interpretation, limited coverage, and low temporal and spatial resolution. While both these methods are powerful, they are often too temporally sparse (i.e., low frequency of sampling) to capture the spatial evolution of VSAs. High temporal resolution measurements of depth to water table are becoming readily available due to inexpensive, self-contained, water level data loggers (e.g., TruTrack, Inc). These loggers can be employed to monitor depth to water table from the field scale up to the watershed scale. The ability to capture short term changes in the water table depths makes it possible to observe the effects of storm events. Since the position of the water table during storm events is crucial to VSAs, these measurement techniques provide new information where the runoff is being produced. However, few techniques are available to summarize this enormous mass of data easily. Here we will show how geostatistics lends itself naturally to characterize spatial patterns and how kriging can be used to interpolate among the points to obtain realistic spatial patterns of water table heights and saturated areas.

Geostatistical analysis most commonly uses semivariograms to define the variance between two observations as a function of the distance separating them. The main parameters of the variogram are the nugget, the sill and the range. The range provides a measure of the maximum distance over which spatial correlation affects the variable of interest. The sill represents the spatial variance of two distant measurements. The nugget represents the variance between two close measurements. The nugget gives the variance in the measurement due the occurrence of spatial patterns smaller than the sampling interval and due to the inherent variability of the sampling device. Within the realm of semivariogram techniques, indicator semivariograms provide a method to capture extreme values (Journel, 1983). Indicator semivariograms have been used to assess risk of contamination in various constituents such as heavy metals (Webster and Oliver, 1989; Smith et al., 1993; Goovaerts and Journel, 1995) and assess uncertainty in soil properties (McKenna, 1998; Pachepsky and Acock, 1998; Goovaerts, 2001). In the most basic form, indicator semivariograms treat data as a binary indicator with respect to a threshold value (i.e., 1 if threshold is exceeded; 0 if threshold is not exceeded). This can be used to identify clustering of extreme values in space. More complete discussions of indicator semivariograms, and the associated kriging, along with many possible derivatives in algorithms and methodology are provided in Goovaerts (1997), Deustch and Journel (1992), and Chilès and Delfiner (1999).

Semivariograms provide the information about the spatial structure needed to interpolate among data points using kriging. Kriging of various forms has been used to interpolate maps of potentiometric surfaces from water table data (Delhomme, 1978; Neuman and Jacobsen, 1984; ASCE, 1990). The goal of most research of this nature is how to best interpolate discrete spatial observations into full coverage. To this end, some research has looked at using existing information about the landscape to supplement point observations of the water table. Hoeksema et al. (1990) supplemented well data with elevation in mapping of a phreatic surface using a cokriging approach. More recently, Desbarats et al. (2002) used kriging with external drift incorporating the TOPMODEL topographic index of Beven and Kirkby (1979) to interpolate water table elevations. Their results showed that predictions made accounting for the traditional topographic index resulted in somewhat non-physical water table behavior in regions of high fluctuations in ground water and sparse observations. Lyon et al. (2005) used indicator kriging (IK) to incorporate soft data developed using logistic regression. "Soft" data are local information that is a proxy to the variable of interest and need not relate directly (Goovaerts, 1997) as opposed to "hard" data which are actual measurements of the variable of interest. They were able to improve interpolations for low antecedent rainfall condition rain events using pre-event water table positions as a predictor of saturation. The analysis of Lyon et al. (2005) required information about the pre-event depth to water table that may not be available and cannot be extended beyond the boundaries of the study site. Also, the study made observations on only large storm events and did not look at how spatial structure changed through time.

This research looked at the spatial and temporal evolution of the shallow water table in the near stream region of a headwater catchment. The position of this shallow water table was directly related to the formation of saturated source areas. Our goal was to characterize both short time interval and long time interval variations, thus better understand event-based and seasonal hydrologic responses, in the spatial structure of the shallow water table using semivariogram analysis. This type of geostatistical analysis is capable of representing large amounts of data easily. Depth to ground water was measured at 44 locations for 5 min intervals from March 2004 through August 2004. These data were used to develop indicator semivariograms at small time intervals over the sampling period using the time-variable median depth to water table as a threshold. By investigating the small time interval changes in characteristics of these indicator semivariograms, the event-based processes driving runoff in the region could be identified. To investigate long time interval changes in the spatial structure of the shallow water table, probability of exceeding the time-variable median depth to water table was computed for each month. This probability was related to the soil topographic wetness index (STWI) for the study site to demonstrate seasonal influences on the spatial structure of the shallow water table. Both the event-based and seasonal influences can be incorporated into a kriging interpolation to visualize the physical patterns occurring in the shallow water table on the hillslope. This geostatistical analysis provides a manner to reinforce spatial observations based on limited, discrete observations using an understanding of the hydrological processes operating on the hillslope. Also, this analysis provides a utility to represent the variability of the shallow water table which affects the formation of saturated regions in both time and space. This representation gives insight to the dominant hydrological patterns in terms of runoff generation at the hillslope scale which can then be located with the help of kriging interpolations. The evolution of these patterns in both space and time directly influences runoff generation and contaminant transport. This makes the correct characterization and representation of them essential for hydrologists interested in predicting water movement from the landscape to the stream.

2. Site description and data

The 2.44 ha study site on New York State Department of Environmental Protection (DEP) owned lands is part of a 2 km^2 sub-watershed located in the southwest corner of the 37 km^2 Townbrook watershed in the Catskill Mountain region of New York State (Fig. 1). The landuse on the study site is uniformly grass/shrub with forested regions upslope (south) the study area. A survey of more than 200 points was conducted to supplement the existing 10 m digital elevation model (DEM) and derive 1-m interval contours for identifying small-scale topographic features. The study site covers the near stream region approximately 120 m along the stream (bordering the northern side of the study site) and 180 m upslope (south) from the stream and elevation varied from 585 m to 600 m above mean sea level with slopes varied from 0° to 8°. Soil Survey Geographic Database (SSURGO) soil maps were used to determine soil types and properties. Two soil types dominate the study site. The northern (down slope) half of the study site consisted of approximately 30 cm deep gravely silt loam. The southern (up slope) half of the study site consisted of approximately 56 cm deep silt loam. The soil is underlain by a restrictive fractured bedrock layer. These shallow soils were typified by a higher hydraulic conductivity (1.4×10^{-5} m/s) in the surface material and a lower hydraulic conductive (1.4×10^{-6} m/s) in deeper layers.

At 44 measurement locations piezometers were instrumented for continuous monitoring depth to water table. The water levels in the upper 30 cm of the soil were recorded using WT-HR 500 capacitance probes manufactured by TruTrack, Inc, New Zealand. Levels were recorded at 5-min intervals and averaged over 15-min intervals for the study period from 10 March 2004 to 22 August 2004. The location of the piezometers followed approximately two grid systems. The first consisted of 20 loggers on a 10×10 m grid near the stream (northern end) of the study site. In addition, 24 loggers were located on a large spacing 30×40 m grid to record water table levels upslope from the stream. A few capacitance probes failed for some periods to record data and need to be repaired, recalibrated, or replaced. During these periods, the sampling location was removed from the data set and assigned a 'no data' value and not used in the analysis. At most, two sampling locations from the 44 sampling locations were assigned 'no data' values at any given time. A tipping bucket rain gauge with data logger was set on the site to record rainfall amounts. Also, two water level loggers were placed in the stream above and below the study site to gauge the runoff from during the sampling period. These water level loggers recorded the stream stage and were converted to flow using rating curves developed for the stream at both locations. Each rating curve was based on seven current-meter discharge measurements. 16% of all stream height observations required extrapolation beyond the highest known point of the rating curve. Runoff from the hillslope was calculated as the difference in flow downstream and upstream of the study site with negative values during low flow conditions removed. This was reasonable since there was only little catchment area contributing from the other side of the stream for this stream segment (Fig. 1). Rain data and stream data were not available for the last two weeks of the study period (from 6 August).

3. Methods

Indicator geostatistics were used to characterize the spatial structure of the shallow water table. To give a proportionate number of observations above and below the threshold, the median depth to water table at each 15-min interval was used as the threshold. Indicator variables were, thus, defined as:

$$I_i(z_c(t)) = \begin{cases} 1 \text{ if } \mathrm{z_i}(t) \leq z_c(t) \\ 0 \text{ if } \mathrm{z_i}(t) > z_c(t) \end{cases} \tag{1}$$

where $I_i(z_c)$ is the indicator value at sampling location i, $\mathrm{z_i}(t)$ is the measured depth to water table at sampling location i [cm] at a certain point in time t, and $z_c(t)$ is the median depth to water table [cm] at the same time t. The time-variable threshold ensured that there were equal numbers of zeros and ones in the data set at any time step. With a constant threshold, the number of ones would be time-variable which would cause artifacts in the geostatistical analysis. It should be noted that a one did not indicate a wet location but rather a location that was wetter than 50% of the wells. The sets of indicator variables for each 15-min time step were used to characterize spatial structure on a short time interval to describe event-based changes in the shallow water table. Long time interval spatial structure at the study site was evaluated by dividing the 15-min data into monthly intervals. Monthly intervals were selected because of their ability to capture the seasonal variability of the spatial distributions of hydrologically active areas for this region (Agnew et al., 2005; Walter et al., 2001). For each month (March through August), the frequency of the water table at a sampling location exceeding the time-variable median water table (i.e., how often was the water table at a certain location among the 50% wettest locations) was computed to give a probability of exceeding the threshold. This frequency also describes the prior probability of exceeding the threshold used later for the development of soft data. Indicator variables were selected because they give information about water table positions deeper than detectable by the piezometers. Also, semivariograms based on indicators may provide additional information over traditional, measurement-based semivariograms for data clustering in space (Western et al., 1998).

Semivariograms were constructed for both the short time interval and the long time interval observations using the semivariance, $\gamma_s(h)$, at a lag, h, of

$$\gamma_s(h) = \frac{1}{2N(h)} \sum_{(i,j)} (Y_i(z) - Y_j(z))^2 \tag{2}$$

where, N is the number of pairs, $Y_i(z)$ and $Y_j(z)$ are the variable of interest at i and j, respectively, with summation over pairs (i, j). For the short time interval, the variable of interest was the indicator values at points i and j instead of measured values. For long time interval, the variable of interest was the probability of exceeding the threshold at i and j. Plotting the average semivariance for pairs grouped by separation distance or grouped into "bins" in semivariogram nomenclature against the average "bin" distance, sample semivariograms were created with Eq. (2) to relate distance between sampling location and semivariance. For the short time interval, the semivariance was normalized with the variance of the observations to lower scatter around the sill. The sample semivariograms were modeled the widely-used exponential relationship (Eq. 3).

$$\gamma_e(h) = \sigma_0^2 + \left(\sigma_\infty^2 - \sigma_0^2\right)\left(1 - e^{\frac{-h}{\lambda}}\right) \tag{3}$$

$\gamma_e(h)$ is the fitted semivariogram, σ_0^2 is the nugget, σ_∞^2 is the sill and λ is the correlation length. This model reaches its sill asymptotically with the range (i.e., maximum distance over which spatial correlation affects the variable of interest) defined as 3λ. Thus, for the short time interval data, indicator semivariograms based on indicator values defined with Eq. (1) were created and for the long time interval data, traditional semivariograms were created from the probability of exceeding the time-variable threshold. Using an automated fitting procedure programmed in Matlab v7r14 (The Mathworks, Inc., 2004) exponential models for both the short time interval indicator semivariograms and long time interval semivariograms were created. Since anisotropy was found to be minimal for the study site (Lyon et al., 2005), only omnidirectional semivariograms were used in this study. The parameters of these models describe the spatial structure of the shallow water table and were compared to measured runoff and surface saturation on the hillslope. For this study, saturation was considered when the depth to water table at a sampling location was less than 5 cm so the water table is close to or at the soil surface. The area representing each sampling location that saturates was determined using Theissen polygons to compute the portion of hillslope saturating.

For this study, two interpolation methods were used to visually identify patterns in the shallow water table on the hillslope. The first was indicator kriging (IK) based on the hard data alone. Using Eq. (1) to create hard data (i.e., indicator variables) from the short time interval data, interpolation between sampling locations (for this and all subsequent interpolations) was made as ordinary kriging performed using the Geostatistical Analyst extension available in ESRI© ArcMap™ v9 (ESRI, Inc., 2004). When using indicator variables, the resulting IK is the probability of exceeding the defined threshold. A major advantage of the IK approach is its ability to account for soft data (Deutsch and Journel, 1992). With this in mind, the second interpolation method for this study was IK coupling hard data with soft data. Soft data can relate prior probabilities about the indicator variables to auxiliary information, such as existing geographic conditions (e.g. soil map, topography). To develop soft data for this study site, the relation between the prior probability (i.e., monthly frequency a sampling location exceeding the time-variable median water table) and the soil topographical wetness index (STWI) was investigated. Agnew et al. (2005) demonstrated that STWI was a good predictor of saturation for this watershed based on a 30-year modeling simulation. STWI is defined as:

$$x_{STWI} = \ln\left(\frac{a}{\tan\beta\ D\hat{K}_s}\right) \tag{4}$$

where a is the area of the upslope watershed per unit contour length [m], tanβ is the local slope, D is the depth of the soil [m] and $\hat{K}_s$ is the mean saturated hydraulic conductivity [m/day]. Values for a and tanβ were determined for the study site using the D_∞ algorithm of Tarboton (1997). D and $\hat{K}_s$ were taken from SSURGO soil distribution maps for the study site. The STWI values from each sampling location were categorized into unit intervals (i.e., sampling locations with STWI values between 8 and 9 in the first category, between 9 and 10 in the next, etc.) and the average STWI was evaluated for each interval. This resulted in six total intervals. The average prior probability for exceeding the median water table was also computed for each interval. A linear function relating STWI to prior probability was then used to create a continuous prior probability map based on STWI or soft data. Residuals were evaluated between the hard data and this map which were interpolated and merged with the soft data using a method consistent with Goovaerts (1997). For comparison, the two interpolation methods were conducted on data from the six rainfall events causing the highest median water tables for the spring period (March through May).

4. Results

4.1. Short time interval

Indicator semivariograms were computed for the study site at each of the 15 min sampling intervals to look at the short time interval spatial structure. The threshold level for the indicator semivariograms was time variable and set at the median depth to water table of all the sampling locations. This provided an equal number of sampling locations above and below the threshold level for any point in time. Exponential models were fitted to the indicator semivariograms for various median water tables and at various times in the sampling period (Fig. 2). The sample indictor semivariograms were calculated using 10 bins with bin sizes of 15 m. Many of the indicator semivariograms had a well defined sill and identifiable ranges. These indicator semivariograms provide information about the spatial structure of the shallow water table for snapshots in time; however, they provide no information about the evolution of the shallow water table with time.

To look at this evolution along with changes in rainfall and runoff at the hillslope, time series were created over the sampling period. Frequent, low intensity storms were more prevalent during the first half of the study period (March through mid-May) while high intensity storms occur less frequently in the second half (Fig. 3A). Peaks in runoff coincided with the rainfall events with large rainfall events producing more runoff from the study site (Fig. 3B). The two largest runoff events occurred after periods of high antecedent rainfall and coincide with large volume rainfall events. The median depth to water table fluctuates quickly, rising in response to rain events for the study site (Fig. 3C). The median water table was consistently close to the ground surface during March through early June. From mid-June through the end of August the median water table was deeper with high fluctuations during rain events. The water table and stream response to rainfall at the study site was typical for this region. High water tables near streams were maintained in spring (March through May) by interflow from either snowmelt or rainfall from upslope areas. The range for the fitted exponential models was highly variable in time (Fig. 3D). The minimum range was about 9 meters and the maximum 105 m.

From this time series, the importance the median water table plays in the hydrology of the hillslope was investigated. The ranges for the short time interval indicator semi-variograms decreased as the median water table rises (Fig. 4A). This trend changes when the median water table was about 10 cm deep. After this point, as the median water table rises closer to the soil surface the ranges began to increase. The runoff increases, as expected, when the median water table rises towards the soil surface (Fig. 4B). There was a large increase in runoff observed when the median water table was closer to the soil surface than 10 cm. There was also an increase in the saturated portion of the hillslope with decrease in median depth to water table (Fig. 4C). For each 2-cm increment in depth to water table the STWI values of all respective piezometers were grouped together and the mean, along with maximum and minimum, were computed. The mean STWI of all the saturated locations decreased as the median water table rose towards the soil surface (Fig. 4D). Also, the minimum STWI for all the sampling locations that saturate tended to decrease as the water table rises to the soil surface while the maximum STWI for all the sampling locations that saturate tends to stay constant.

4.2. Long time interval

Monthly intervals were selected to characterize the long time interval spatial structure in the shallow water table. This interval sufficiently captures the seasonal variability of the spatial distributions of hydrologically active areas (Agnew et al., 2005; Walter et al., 2001). For each month (March through August), the frequency of the water table at a sampling location exceeding the time-variable median water table was computed and used to develop semivariograms. The semivariograms for the long time interval data all had well defined sills and ranges (Fig. 5, Table 1). The nugget values for all months were similar ranging from 0.036 in August to 0.54 in March and April. Due to uncertainty associated with these nuggets, no further conclusions could be drawn from the nugget values. Sills, which are representative of the variance of the measures, varied from higher values during low median depth to water tables (0.193 in April, and 0.194 in May) to lower values during high median depth to water tables (0.147 in June and 0.135 in July). The ranges for the exponential models were longest in June and July at 73.8 and 144.4 m, respectively, and shorter during months with low median depth to water table.

The long time interval analysis provided a prior probability of exceeding the threshold on a monthly basis. This prior probability representing seasonal variability, in turn, defined soft data capable of being incorporated with hard data using IK techniques to visualize variations in spatial patterns of the shallow water table. To create this soft data, the probability of exceeding the median water table was correlated to STWI for March through August (Fig. 6). There was a noticeable difference in the relation between prior probability and STWI when comparing March through May and June though August. For March through May, the low depth to median water table due to spring thaw caused increased prior probability with increased STWI with a higher slope in the linear regression equation. For June through July, the slope of the linear regression equation was much lower. R^2 values were low for July and August.

The combined influence of long time interval and short time interval information on the spatial structure of the shallow water table was demonstrated visually for six rain fall events using kriging techniques (Fig. 7). These events were selected because they produced the highest median water tables for the period from March through May (i.e., when there was a noticeable increase in probability with increase in STWI) and characterize the hillslope response to rainfall for wet conditions (Table 2). For the 27 March and 3 May events, IK interpolations based on hard data alone showed high probability of exceeding the median water table in the near stream region. Also, there was a region of high probability extending up the hillslope. Within this upslope region, there occurred discontinuous islands of higher probabilities. Incorporating soft data based on seasonal trends in the spatial structure of the water table into the IK interpolation reduced the occurrence of these isolated islands of high probabilities. For the events on 2, 13, and 26 April and 26 May, IK interpolations based on hard data alone gave relatively high probability of exceeding the threshold for the area closest to the stream, but little high probability in the region further upslope. Incorporating soft data, the near stream region having high probability of exceeding the median water table was larger. Also, the topographically converging region upslope from the stream was predicted as having higher probability of exceeding the threshold when incorporating the soft data than when using hard data alone.

To quantify the improvement in interpolation made by incorporating soft data, jack-knifing was used to cross validate the kriging interpolations. This method of cross validation tests a kriging interpolation by dividing the original dataset to produce a testing and a training dataset. Randomly, 30% (14 of the 44 total) of the sampling locations were removed from the original dataset to create a testing dataset leaving 70% (30 of the 44 total) in a training dataset for analysis. To compare the interpolation methods, root mean square error (RMSE) was computed between the observed values in the testing dataset and predicted values using both methods. From these, the percentage reduction achieved by incorporating soft data evaluated. For each event, IK incorporating soft data reduces the RMSE for between the observed and predicted values (Table 2). This reduction in RMSE reflects a better spatial representation of observed depths to water table using IK with soft data.

5. Discussion

Semivariograms based on probability of exceeding the median water table or the long time interval analysis (Fig. 5) have a clear distinction between wet conditions (March, April, and May) and dry conditions (June, July, August). This demonstrates the seasonal controls on hydrology for this hillslope. During wet, spring conditions, shallow water tables in the convergent zones lead to shorter ranges in the long time interval semivariogram results. This is similar to the results of Western et al. (1998b) for soil moisture distributions in the Tarrawarra watershed. Locations where the water table is likely to rise during rainfall events are spatially closer together during wet conditions than during dry. This agrees with the short time interval results where saturation causes increased ranges in indictor semivariograms. Since the expansion and contraction of saturated regions occurs quickly, the longer ranges seen during rain events in the short time interval analysis are not reflected in the monthly, frequency-based analysis. The long time interval analysis is influenced by seasonal variations in shallow water table while the short time interval analysis is influenced by the formation of surface saturated areas due to individual rainfall events. When the median water table is close to the surface, topographic converging areas concentrate interflow and produce regions of higher water table. The higher water tables are prerequisite for the lateral expansion of large-scale saturated source areas seen in the short time interval analysis. Previous work showed a correlation between pre-event depth to water table and probability of saturation at the sampling locations during wet conditions (Lyon et al., 2005). During the dry, summer period of the study (June, July, and August), there is a reduction in variance and an increase in range for the long time interval analysis. The water table is more uniform in space on average for these months. Locations where the water table is likely to rise in response to rainfall are spatially distributed across the hillslope. Thus, the relation between STWI and prior probability (Fig. 4) shifts for these summer months. The long term analysis provides more information about prior conditions for the hillslope which describe the seasonal change in water table response as we move from the wet period to the dry period.

The short time interval analysis provides a way to describe changes in the spatial structure of the shallow water table in response to rainfall events which, for this study site, are influenced by the antecedent conditions. During wet conditions, the local water table is close to the soil surface between rainfall events and, when rainfall occurs, the median water table raises producing surface saturation (Fig. 4C). The lateral extent of expansion is captured with the decreasing minimum (and constant maximum) STWI for these saturated areas as the median water table approaches the soil surface (Fig. 4D). Since these saturated regions expand along gradients of decreasing STWI, the indicator semivariograms exhibit increased ranges for these rainfall events producing large, expanding saturated areas. This reaction is common to watersheds where saturation excess overland flow is considered a dominant pathway during wet season rain events (Western et al., 2004). This lateral expansion of saturated areas has been observed by other researchers in the Catskills due to accumulation of interflow water in the form of increased soil moisture at the hill bottoms relative to the steep parts of the hills during wet periods (Frankenberger et al., 1999; Ogden and Watts, 2000; Mehta et al., 2004). Some observed locations where saturation commonly occurs are those where (1) the soil above the low conductivity layer is shallow, (2) the slope decreases downhill, such as the toe-slope of a hill, or (3) in topographically converging areas. In this study, occurrences of exceeding the median water table, which my be an adequate surrogate for saturation during high water table conditions, was observed at all three locations. Shallower soils and a toe-slope occur in the region adjacent the stream while a topographical convergence occurs in the upper hillslope.

Without additional information such as provided by environmental traces it is not possible to discern the exact hydrological pathways. Still, identifying spatial patterns of saturation is assumed to provide important information when focusing on topics such as non-point source pollution control (Walter et al., 2005). Throughout the observation period, non-linear increasing runoff was observed when there were increasing saturated areas (Fig. 4A). A possible interpretation is that as surface saturation regions expand, more rainfall is directly contributing to stream flow as overland flow. There seemed to be a threshold above which the median water table must rise before runoff to the stream increased dramatically (Fig. 4B). It is when the median water table raises above this threshold that longer ranges are observed in the indicator semivariograms due to expansion of surface saturated areas. The identification of these spatial patterns of saturation, which can be used to control where and when chemicals and nutrients can be applied, can be an important hydrological component for non-point source pollution control. Using kriging techniques (Fig. 7), the semivariogram analysis used to investigate the spatial and temporal evolution of the shallow water table can be further employed to identify such physical patterns on the hillslope.

The use of geostatistical techniques, such as IK, is influenced by the number of sampling locations. Western et al. (1998a) suggest that a large dataset is required to produce reliable results. For this study, since the water table was below the extent of the sampling devices over parts of the sampling period, traditional, measurement-based semivariogram analysis was not an option. By transforming measures into indicator values, water tables deeper than the sampling devices could still be included in the analysis. The limited number of sampling locations can produce large fluctuations in the indicator semivariogram ranges for the short time interval analysis. This can lead to poor representations when kriging. However, the length of the sampling period has allowed for the use of soft data in combination with IK to create a more robust interpolation of the observed data that incorporates different timescales. This compensates for sparse spatial coverage and incorporates the seasonal variations in the hydrology of the region into the dataset. The STWI used in this study correlated well with probability of exceeding the median water table during the wet period. This is similar to the long-term modeling results seen for this watershed by Agnew et al. (2005) where the probability of saturation developed from a 30 years modeling study correlated well to STWI. When the median water table is close to the soil surface, such as in periods of snow melt, the probability of exceeding the median water table coincides to the probability of saturation. The influence of topography during drier periods when the water table is not near the soil's surface, however, is not well established. For the wet period, the soft data created with prior probability allowed for IK that represents the physical process of the hillslope. This smoothed the kriging and eliminated islands of high probability of exceeding the threshold. These isolated regions are attributed to the sparse nature of the point observations and position of sampling sites influencing the IK. Using soft data, the data are interpolated in a manner consistent with the underlying hydrologic processes for the hillslope to represent the influence of both event-based and seasonal trends.

By incorporating the soft data with the IK, the overall error in interpolation for the data was reduced. This provides better information about where on the hillslope hydrologically active areas occur. These regions are extremely important in the development of nutrient management plans and in control the transport of pollutants. Also, by developing soft data based on readily available spatial data (i.e., DEM), prior probabilities could be developed from other analysis techniques if long time interval data such as those used in this study are not available. Lyon et al. (2005) improved interpolations by incorporating soft data into IK, but these improvements are limited to locations where the pre-event water table is known explicitly. Long-term modeling studies, such as that of Agnew et al. (2005), can provide the prior probability to create soft data for fortifying hard data observations. Thus, fewer observations can be made without compromising the robustness of the spatial data obtained. In addition, this soft data, when occurring at a longer temporal scale, can provide information about seasonal variations in spatial patterns that heavily influence hydrology. Data based on interpolations of this style provide sources for validation of long-term risk assessment models. They can also aid in the development of appropriate techniques to better model saturated area formation by spatially representing data about water table response to rainfall events. Incorporation of soft data leads to a more realistic representation of hillslope reaction to rainfall events by including processes involved in the formation of saturated areas. This style of geostatistical analysis gives a manner to organize and represent spatial changes in the shallow ground water table. These changes occur at different temporal scales that can be integrated to better describe physical hydrology at the hillslope scale.

6. Conclusions

Geostatistical methods have been used to describe the spatial structure of the shallow water table in the near stream region. Using 44 sampling locations from a study site in the Townbrook watershed in the Catskill Mountain region of New York State, indicator methods have been used to explore variations in both short time intervals (15-min) and long time intervals (months). These time intervals were able to describe the event-based reaction of the shallow water table and the seasonal trends influence the hydrology of the hillslope. The shallow water table for the study site shows two distinct responses depending on the position of the median water table. When the median water table was near to the soil surface (wet conditions) the variability in the depth of the shallow water table decreased with rain resulting in longer ranges in the indicator semivariograms. This is caused by expansion of saturated areas in topographically converging zones. During dry periods when the median water table was far from the soil surface, however, the shallow water table showed more spatial homogeneity prior to rainfall events. It was possible to represent these changes in spatial structure using kriging techniques incorporating both the event-based and seasonal trends in the shallow water table response. From the long time interval, seasonal variations in spatial structure, prior probability at each sampling location was established to create soft data using STWI. The soft data was combined with hard data from the short time interval, event-based variations using IK techniques. This provided more realistic interpolations during high water table conditions by capturing structure in the shallow water table not available when using hard data alone. This type of kriging analysis provides a manner to locate physical patterns influencing the hydrology of the study site. This study presents methods to characterize large amounts of point data temporally and spatially that can emphasize the physical hydrology of a field site. By representing both spatial patterns and temporal evolution in the shallow water table with geostatistical analysis, saturated source areas active in controlling not only VSA runoff but also other hydrological pathways can be identified. Understanding this temporal evolution in the spatial structure of the shallow water table is the ‘where’ and ‘when’ of hydrology that is the groundwork for tasks such as non-point source pollution control.

Acknowledgements. Research is made possible with partial funding from the Department of Interior, US Geological Survey and the Cornell University, New York Resources Institute under Grant Agreement No 01HQGR0208 and with partial funding from the Swedish Research Council (grants 620-20001065/2001). In addition, the first author would like to thank the American-Scandinavian Foundation for funding that made possible the collaboration between researchers at Cornell University and Stockholm University. Finally, special thanks go to the New York City

Department of Environmental Protection for making their lands available for the field work.

References

Agnew, L. J., Lyon, S. W., Gérard-Marchant, P., Collins, V. B., Lembo, A. J., Steenhuis, T. S., and Walter, M. T.: Identifying hydrologically sensitive areas: Bridging the gap between science and application, Hydrol. Processes, in press, 2005.

ASCE American Society of Civil Engineers Task Committee on geostatistical techniques in geohydrology: Review of geostatistics in geohydrology 1: Basic concepts; 2: Applications, ASCE J. Hydrualic Eng., 116, 5, 612–658, 1990.

Beven, K. J. and Kirkby, M. J.: A physically based, variable contributing area model of basin hydrology, Hydrol. Sci. Bull., 24, 3–69, 1979.

Blyth, K.: The use of microwave remote sensing to improve spatial parameterization of hydrological models, J. Hydrol., 152, 103–129, 1993.

Chilès, J.-P. and Delfiner, P.: Geostatistics: Modeling spatial uncertainty, Wiley, New York, 1999.

Choudhury, B. J.: Multispectral satellite data in the context of land surface heat balance, Rev. Geophys., 29, 2, 217–236, 1991.

Delhomme, J. P.: Kriging in the hydrosciences, Ad. Wat. Resour., 1, 5, 251–266, 1978.

Desbarats, A. J., Logan, C. E., Hinton, M. J., and Sharpe, D. R.: On the kriging of water table elevations using collateral information from a digital elevation model, J. Hydrol., 255, 25–38, 2002.

Deutsch, C. V. and Journel, A. G.: GSLIB: Geostatistical software library and user's guide, Oxford University Press, New York, 1992.

Dunne, T. and Black, R. D.: Partial area contributions to storm runoff in a small New-England watershed, Wat. Resour. Res., 6, 5, 1296–1308, 1970.

Dunne, T.: Runoff production in humid areas, U.S. Depart. of Agric. Publication, ARS-41-160, 108 pp., 1970.

Engman, E. T.: Recent advances in remote sensing in hydrology, U.S. Natl. Rep. Int. Union Geod. Geophys. 1991–1994, Rev. Geophys., 33, 967–975, 1995.

Frankenberger, J. R., Brooks, E. S., Walter, M. T., Walter, M. F., and Steenhuis, T. S.: A GIS-based variable source area model, Hydrol. Processes, 13, 804–822, 1999.

Goovaerts, P. and Journel, A. G.: Integrating soil map information in modeling the spatial variation of continuous soil properties, Eur. J. Soil Sci., 46, 397–414, 1995.
Goovaerts, P.: Geostatistics in natural resources evaluations, Oxford University Press, New York, 1997.
Goovaerts, P.: Geostatistical modeling of uncertainty in soil science, Geoderma, 103, 3–29, 2001.
Grayson, R. B., Blöschl, G., Western, A. W., and McMahon, T. A.: Advances in the use of observed spatial patterns of catchment hydrological response, Adv. Wat. Resour. Res., 25, 1313–1334, 2002.
Hewlett, J. D. and Hibbert, A. R.: Factors affecting the response of small watersheds to precipitation in humid regions, in: Forest Hydrology, edited by: Sopper, W. E. and Lull, H. W., Pergamon Press, Oxford, 275–290, 1967.
Hillel, D.: Modeling in soil physics: A critical review, in: Future developments in Soil Science Research, A collection of Soil Science Society of America Golden Anniversary contributions presented at Annual Meeting, New Orleans, 35–42, 1986.
Hoeksema, R. J., Clapp, R. B., Thomas, A. L., Hunley, A. L., Farrow, N. D., and Dearstone, K. C.: Cokriging model for estimation of water table elevation, Wat. Resour. Res., 25, 3, 429–438, 1989.
Hornberger, G. M. and Boyer, E. W.: Recent advances in watershed modeling, U.S. Natl. Rep. Int. Union Geod. Geophys. 1991–1994, Rev. Geophys., 33, 949–957, 1995.
Journel, A. G.: Nonparametric estimation of spatial distributions, Math. Geol., 15, 445–468, 1983.
Klemeš, V.: Dilettantism in hydrology: transition or destiny, Wat. Resour. Res., 22, 177–188, 1986.
Lehmann, W.: Anwendung geostatistischer verfahren auf die bodenfeuchte in läandlichen einzugsgebieten, PhD thesis, Universität Karlsruhle, Karlsruhle, 1995.
Lyon, S. W., Lembo, A. J., Walter, M. T., and Steenhuis T. S.: Defining probability of saturation with indicator kriging on hard and soft data, Adv. Wat. Resour., in press, 2005.
McDonnell, J. J.: Where does water go when it rains? Moving beyond the variable source area concept of rainfall-runoff response, Hydrol. Processes, 17, 1869–1875, 2003.
McKenna, S. A.: Geostatistical approach for managing uncertainty in environmental remediation of contaminated soils: case study, Environ. Eng. Geosci., 4, 175–184, 1998.
Mehta, V. K., Walter, M. T., Brooks, E. S., Steenhuis, T. S., Walter, M. F., Johnson, M., Boll, J.,

and Thongs, D.: Evaluation and application of SMR for watershed modeling in the Catskill Mountains of New York State, Environ. Modeling and Assessment, 9, 2, 77–89, 2004.
Meyles, E., Williams, A., Dowd, J., and Ternan, L.: Soil moisture patterns and runoff generation in a small Dartmoor catchment, south-west England, in: Leibundgut Ch, edited by: Uhlenbrook, S. and McDonnell, J., Runoff generation and implications for river basin modeling, Freiburger Schriften Hydrologie, band 13, Institute of Hydrology, University of Freiburg i.Br./Germany, 28–36, 2001.
Mohanty, B. P., Skaggs, T. H., and Famiglietti, J. S.: Analysis and mapping of field-scale soil moisture variability using high-resolution, ground-based data during the Southern Great Plains 1997 (SGP97) hydrology experiment, Wat. Resour. Res., 36, 4, 1023–1032, 2000.
Neuman, S. P. and Jacobsen, E. A.: Analysis of non-intrinsic spatial variability by residual kriging with applications to regional ground water levels, Math. Geo., 16, 5, 499–521, 1984.
Ogden, F. L. and Watts, B. A.: Saturated area formation on nonconvergent hillslope topography with shallow soils: A numerical investigation, Wat. Resour. Res., 36, 7, 1795–1804, 2000.
Pachepsky, Y. and Acock, B.: Stochastic imaging of soil properties to assess variability and uncertainty of crop yield estimates, Geoderma, 85, 213–229, 1998.
Smith, J. L., Halvorson, J. J., and Papendick, R. I.: Using multiple-variable indicator kriging for evaluating soil quality, Soil Sci. Soc. Am. J., 57, 743–749, 1993.
Tarboton, D. G.: A new method for the determination of flow directions and contributing areas in grid digital elevation models, Wat. Resour. Res., 33, 2, 309–319, 1997.
Troch, P., Verhoest, N., Gineste, P., Paniconi, C., and Merot, P.: Variable source areas, soil moisture, and active microwave observations at Zwalmbeek and Coët-Dan, in: Spatial patterns in catchment hydrology: observations and modeling, edited by: Grayson, R. and Blöschl, G., Cambridge University Press, Cambridge, 8, 187–208, 2000.
Tyndale-Biscoe, J. P., Moore, G. A., and Western, A. W.: A system for collecting spatially variable terrain data, Comput. Electron. Agric., 19, 113–128, 1998.
Verhoest, N. E. C., Paniconi, C., and De Troch, F. P.: Mapping basin scale variable source areas from multitemporal remotely sensed observations of soil moisture behavior, Wat. Resour. Res., 24, 12, 3235–3244, 1998.
Walker, J. P., Willgoose, G. R., and Kalma, J. D.: The Nerrigundah data set: soil moisture patterns, soil characteristics, and hydrological flux measurements, Wat. Resour. Res., 37, 11, 2653–2658, 2001.
Walter, M. T., Brooks, E. S., Walter, M. F., Steenhuis, T. S., Scott, C. A., and Boll, J.: Eval-

uation of soluble phosphorus loading from manure-applied fields under various spreading strategies, J. Soil Wat. Conserv., 56, 329–335, 2001.
Walter, M. T., Gérard-Marchant, P., Steenhuis, T. S., and Walter, M. F.: Closure to "Simple Estimation of prevalence of hortonian flow in New York City watersheds" by M. Todd Walter, Vishal K. Mehta, Alexis M. Marrone, Jan Boll, Pierre Gérard-Marchant, Tammo Steenhuis, and Michael F. Walter, ASCE J. Hydrol. Eng., 10, 2, 169–170, 2005.
Webster, R. and Oliver, M. A.: Optimal interpolation and isrithmic mapping of soil properties: VI. Disjunctive kriging and mapping the conditional probability, J. Soil Sci., 40, 497–512, 1989.
Western, A. W. and Grayson, R. B.: The tarrawarra data set: soil moisture patterns, soil characteristics, and hydrological flux measurements, Wat. Resour. Res., 34, 10, 2765–2768, 1998.
Western, A. W., Blöschl, G., and Grayson, R. B.: How well do indicator variograms capture the spatial connectivity of soil moisture?, Hydrol. Processes, 12, 1851–1868, 1998a.
Western, A. W., Blöschl, G., and Grayson, R. B.: Geostatistical characterisation of soil moisture patterns in the Tarrawarra catchment, J. Hydrol., 205, 20–37, 1998b.
Wilson, D. J., Western, A. W., and Grayson, R. B.: Identifying and quantifying sources of variability in temporal and spatial soil moisture observations, Wat. Resour. Res., 40, W02507, doi:10.1029/2003WR002306, 2004.

Table 1. Monthly median depth to water table, total rainfall, and semivariogram parameters for the exponential models in Fig. 4 using the long time interval data for the study site.

Month	Depth to water table [cm]	Total rainfall [cm]	Semivariogram Parameters		
			Nugget	Sill	Range [m]
March	16.6	6.1	0.054	0.173	12.0
April	13.05	10.2	0.054	0.193	17.3
May	15.75	18.3	0.048	0.194	20.9
June	31.85	8.7	0.049	0.147	73.8
July	35.25	15.2	0.039	0.135	144.4
August	16.51	12.2	0.036	0.161	29.0

Table 2. Summary of water table and rainfall for six dates used in IK analysis along with reduction in RMSE from cross validation with jackknifing between IK with hard data alone and IK with soft data.

Date	Median depth to water table [cm]	20th/80th percentile depth to water table [cm]	1 day antecedent rainfall [cm]	Reduction in RMSE (%)
27 March	8.4	2.9/16.6	1.1	4.3
2 April	6.2	1.3/12.8	1.5	11.7
13 April	7.7	2.6/19.2	2.7	8.5
26 April	5.8	1.0/12.7	2.4	1.2
3 May	6.4	0.5/17.2	2.1	9.9
26 May	4.4	0.4/10.5	3.6	7.9

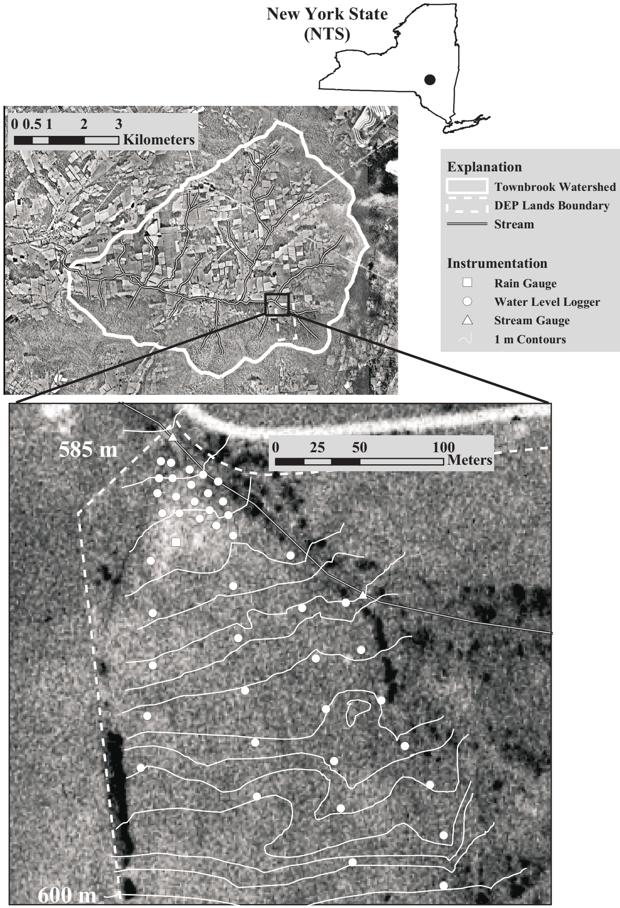

Fig. 1. Location of study site at Townbrook research watershed in Catskill Mountain region of New York State showing positions of sampling locations containing water level loggers (circles), stream gauges (triangles), and rain gauge (square).

HESSD

2, 1683–1716, 2005

Spatial structure in shallow water table

S. W. Lyon et al.

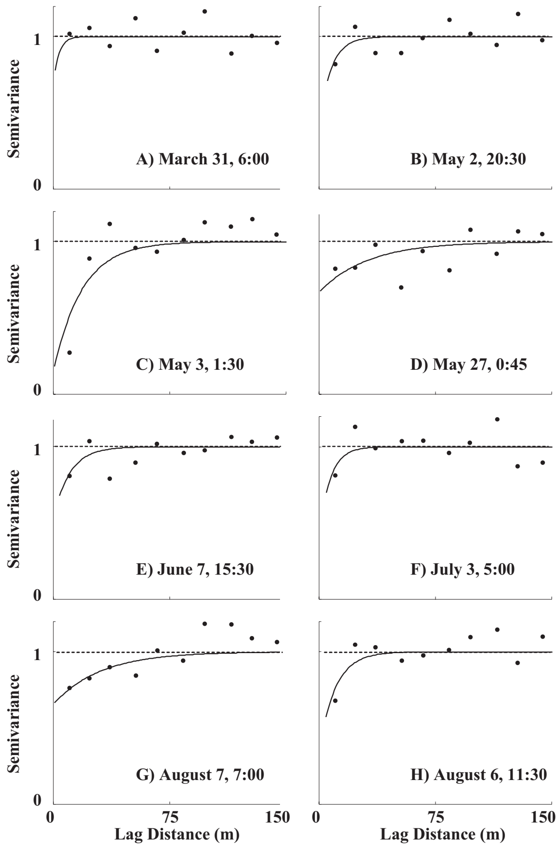

Fig. 2. Typical indicator semivariograms from the short time interval data for the study site for **(A)** 31 March, 06:00, **(B)** 2 May, 20:30, **(C)** 3 May, 01:30, **(D)** 27 May, 00:45, **(E)** 7 June, 15:30, 30 June, 04:30, **(G)** 26 July, 05:30, and **(H)** 6 August, 11:30. The symbols are the normalized sample indicator semivariogram and the curves are the fitted exponential models.

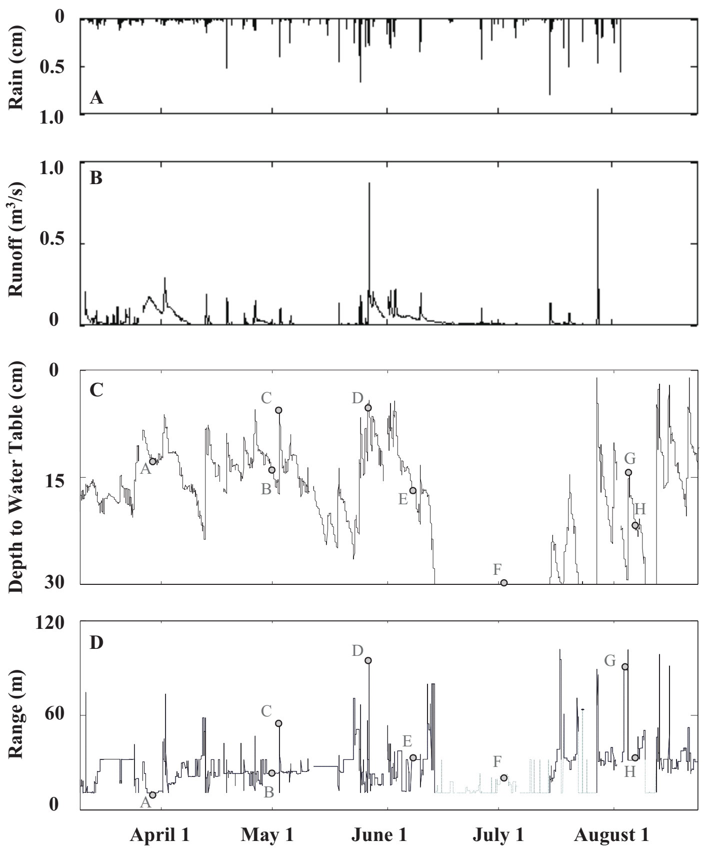

Fig. 3. Short time interval measurements for the study site from 10 March through 22 August of **(A)** rainfall [cm], **(B)** runoff [m^3/s], and **(C)** median depth to water table [cm]. For each indicator semivariogram, the **(D)** range parameter from exponential model [m]. Circles on (C) and (D) show when in time the indicator semivariograms in Fig. 2 occur and the portion of (D) indicated with a dashed line is where number of sampling sites below minimum detection level is greater than half all sampling sites.

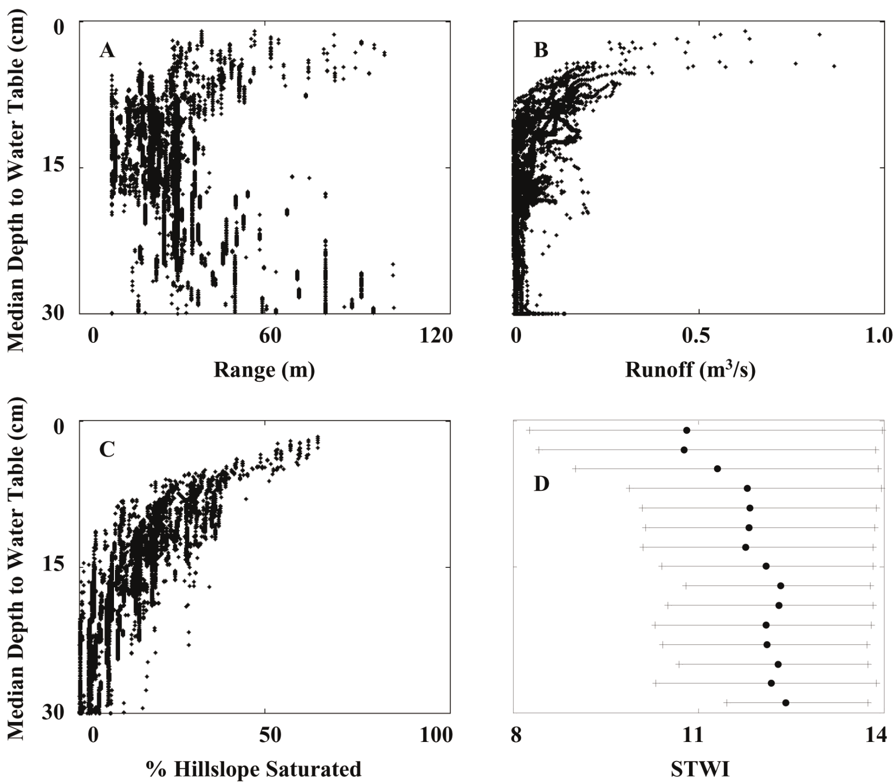

Fig. 4. From the short time interval data, variations in **(A)** the range [m], **(B)** runoff [m^3/s], **(C)** percentage of the hillslope saturated [%], and **(D)** average (black dots) STWI of the saturated area with bars showing minimum and maximum STWI of the saturated area with respect to the median depth to water table [cm] for the study site.

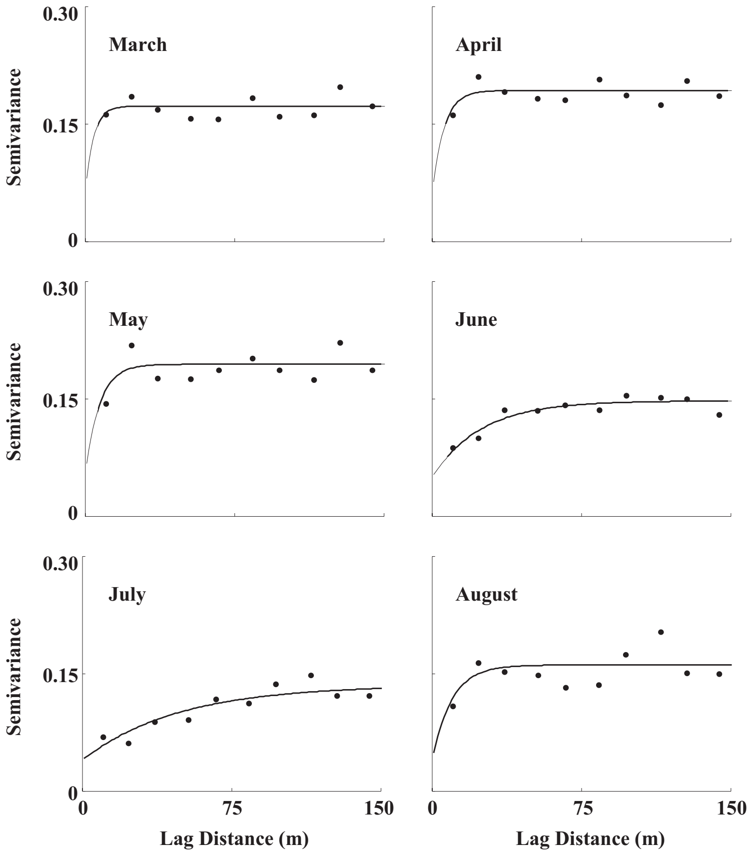

Fig. 5. Semivariograms for study site using the probability of exceeding the time-variable threshold (median water level) at each sampling location during the different months using long time interval analysis. The symbols are the sample semivariogram and the curves are the fitted exponential models.

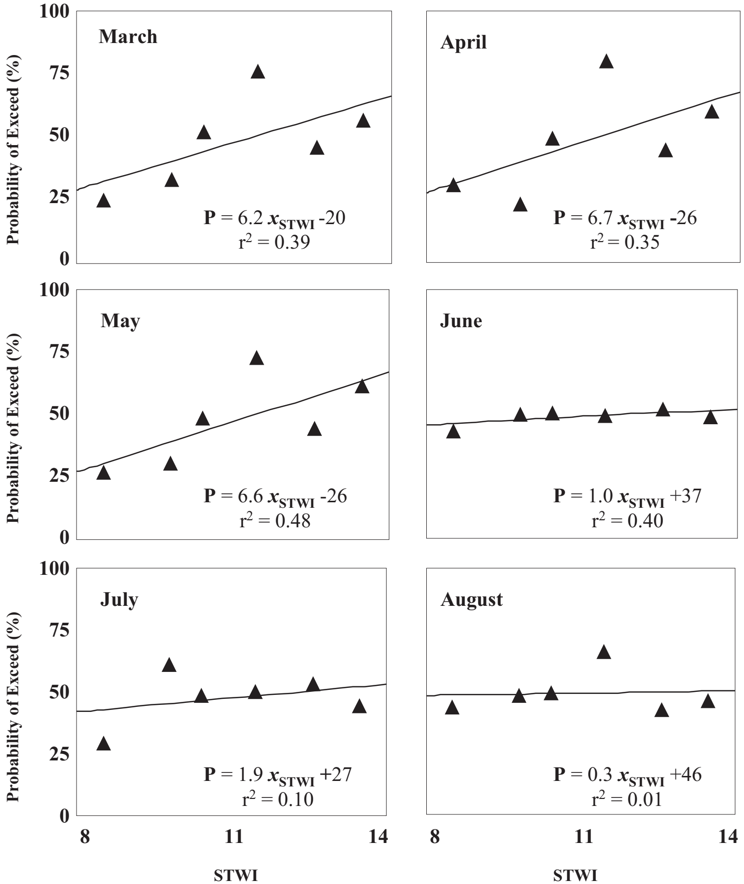

Fig. 6. Relationship between probability of exceeding the time-variable threshold (median water level) and STWI for each month. Points represent average probability of exceeding threshold and average STWI for each unit interval of STWI for the hillslope.

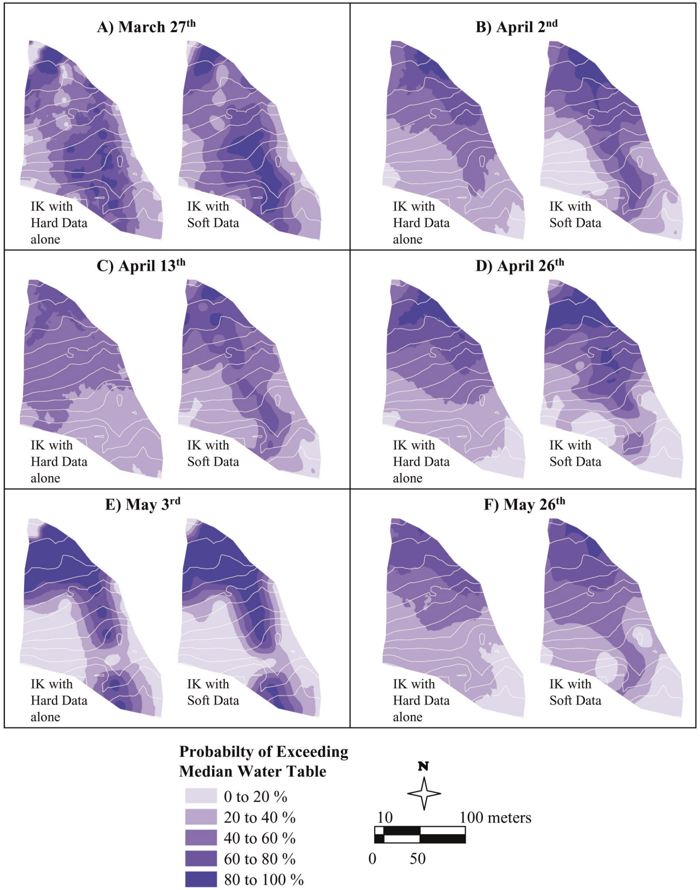

Fig. 7. IK with hard data alone and IK with soft data of study site for **(A)** 27 March, **(B)** 2 April, **(C)** 13 April, **(D)** 26 April, **(E)** 3 May, and **(F)** 26 May rain events using indicator values from short time interval for peak in rise of water table with 1 m contours as white lines.